UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/487,191	01/19/2000	Rakesh Agrawal	AM9-99-0226	2881

7590 11/06/2002

John L Rogitz
Rogitz & Associates
Suite 3120
750 B Street
San Diego, CA 92101

EXAMINER	
FLEURANTIN, JEAN B	
ART UNIT	PAPER NUMBER
2172	

DATE MAILED: 11/06/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Paper No. 8

Application Number: 09/487,191
Filing Date: January 19, 2000
Appellant(s): AGRAWAL ET AL.

Rakesh Agrawal et al.
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed August 15, 2002.

(1) Real Party in Interest

A statement identifying the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

A statement identifying the related appeals and interferences which will directly affect or be directly affected by or have a bearing on the decision in the pending appeal is contained in the brief.

(3) Status of Claims

The statement of the status of the claims contained in the brief is correct.

(4) Status of Amendments After Final

No amendment after final has been filed.

(5) Summary of Invention

The summary of invention contained in the brief is correct.

(6) Issues

The appellant's statement of the issues in the brief is correct.

(7) Grouping of Claims

Appellant's brief includes a statement that the claims stand and fall together.

(8) Claims Appealed

The copy of the appealed claims contained in the Appendix to the brief is correct.

(9) Prior Art of Record

The following is a listing of the prior art of record relied upon in the rejection of claims under appeal.

6,115,708	Fayyad et al.	09-2000
A Modified Random Perturbation Method for Database Security	Tendick et al.	03-1994

(10) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims 1, 7 and 13:

Claims 1-13 and 20-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Fayyad et al. (US Pat. No. 6,115,708) in view of Tendick et al. 'A Modified Random Perturbation Method for Database Security - 03/1994' ("Fayyad"), ("Tendick"). This rejection is set forth in prior Office Action, mailed on July 15, 2002, Paper No. 5.

I.

Claim Rejections - 35 U.S.C. § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-13 and 20-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Fayyad et al. (US Pat. No. 6,115,708) in view of Tendick et al. 'A Modified Random Perturbation Method for Database Security - 03/1994' ("Fayyad"), ("Tendick").

As per claim 1, Fayyad substantially teaches the steps of perturbing original data associated with the user computer to render perturbed data (thus, some methods take the mean of the global data set and perturb it K times to get the K initial means or simply pick K random

points from the data set, in most situations initialization is done by randomly picking a set of starting points from the range of the data; which is readable as perturbing original data associated with the user computer to render perturbed data) (see col. 2, lines 32-36);

using a distribution of the perturbed data, generating at least on estimate of a distribution of the original data; (thus, the variance in result illustrated by these depictions is fairly common even in low dimensions using data from well-separated Gaussians, these figures also illustrate the importance of the problem of having a good initial or starting point each of the two data clusters depicted in FIGS. 5A and 5B depict clustering from 2 different samples of the same size that were obtained from the same database; which is readable as using a distribution of the perturbed data, generating at least on estimate of a distribution of the original data) (see col. 6, lines 25-32); and

using the estimate of the distribution of the original data, generating at least one data mining model (thus, the end of the clustering process any of the clusters have zero membership then the corresponding initial guess at this cluster centroid is set to the data point farthest from its assigned cluster center, this procedure decreases the likelihood of having empty clusters after reclustering from the "new" initial point, resetting the empty centroids to another point may be done in a variety of ways; which is readable as using the estimate of the distribution of the original data, generating at least one data mining model) (see col. 7, lines 32-39). But, Fayyad does not explicitly indicate the step of maintaining the privacy of a user of the computer as claimed in the preamble. However, Tendick implicitly indicates steps of the database management system must include mechanisms which allow statistical analysis but not access to data individual database records, which is readable as maintaining the privacy of a user of the computer (see page 48, lines 6-7). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Fayyad and Tendick with steps of maintaining the privacy of a user of the computer. This modification would allow the

teachings of Fayyad and Tendick to improve the performance of the system and architecture for privacy preserving data mining, and provide the optimal protection against such a problem among all possible covariance structures, and a certain specified level of statistical accuracy for legitimate users of the database (see page 61, lines 4-7).

As per claims 2 and 8, Fayyad substantially teaches a method as claimed, wherein perturbed data is generated from plural original data associated with respective plural user computers (thus, the end of the clustering process any of the clusters have zero membership then the corresponding initial guess at this cluster centroid is set to the data point farthest from its assigned cluster center, this procedure decreases the likelihood of having empty clusters after reclustering from the "new" initial point, resetting the empty centroids to another point may be done in a variety of ways; which is readable as wherein perturbed data is generated from plural original data associated with respective plural user computers) (see col. 7, lines 32-39).

As per claims 3 and 9, Fayyad wherein the original values cannot be reconstructed from the respective perturbed values (thus, if at the end of the clustering process any of the clusters have zero membership then the corresponding initial guess at this cluster centroid is set to the data point farthest from its assigned cluster center this procedure decreases the likehood of having empty clusters after reclustering from the new initial point resetting the centroids to another point may be done in a variety of ways, which is readable as wherein the original values cannot be reconstructed from the respective perturbed values) (see col. 7, lines 37-39).

As per claims 4 and 10, Fayyad substantially teaches a method as claimed, wherein at least some of the data is perturbed using a uniform probability distribution (thus, the result of clustering two different subsamples drawn from the same distribution and initialized with the same starting point, which is readable as wherein at least some of the data is perturbed using a uniform probability distribution) (see col. 6, lines 22-26).

As per claims 5 and 11, Fayyad substantially teaches a method as claimed, wherein at least some of the data is perturbed using a Gaussian probability (thus, the model cluster is assumed be a Gaussian for each cluster the Gaussian is centered at the mean of the cluster, which is readable as wherein at least some of the data is perturbed using a Gaussian probability) (see cols. 2-3, lines 67-2).

As per claim 6, Fayyad substantially teaches a method as claimed, wherein at least some of the data is perturbed by selectively replacing the data with other values based on a probability (thus, a multinomial distribution has a simple set of parameters for every attribute a vector of probabilities specified the probabilities of each value of the attribute given the cluster, which is readable as wherein at least some of the data is perturbed by selectively replacing the data with other values based on a probability) (see col. 10, lines 20-25).

As per claims 7 and 13, in addition to the discussion in claim 1, Fayyad further teaches steps of sending the perturbed values to a server computer not having access to the original values (thus, data samples are each chosen as a starting point for a clustering of all the candidate solutions, the best solution returned as the refined 'improved' starting point to be used in clustering the full data set; which is readable as sending the perturbed values to a server computer not having access to the original values) (see col. 3, lines 37-41). Also, in column 11, lines 36 through 40, Fayyad further teaches steps of finding multiple candidate clustering starting points from the multiple data subsets retrieved from the database and choosing an optimum solution from the multiple number of candidate clustering starting points to begin subsequent clustering on data in the database.

As per claim 12, Fayyad substantially teaches a method as claimed, wherein the method acts further comprise perturbing categorical values of at least some categorical attributes by selectively replacing the categorical values with other values based on a probability (thus, assume the user is clustering using the EM algorithm and that data is discrete, and hence each

cluster specifies a multinomial distribution over the data a multinomial distribution has a simple set of parameters for every attribute a vector of probabilities specified the probabilities of each value of the attribute given the cluster, since these probabilities are continuous quantities they have a "centroid" and K-means can be applied to them; which is readable as wherein the method acts further comprise perturbing categorical values of at least some categorical attributes by selectively replacing the categorical values with other values based on a probability) (see col. 10, lines 18-25).

As per claims 20 and 23, Fayyad substantially teaches a method as claimed, further comprises step of sending the model to at least one user computer for use thereof by the user computer on original data (thus, a mixture model M having K clusters Ci, I=1 . . . , K assigns a probability to a data point x as follows ##EQU1## where W.sub.i are called the mixture weights, the problem of clustering is identifying the properties of the clusters Ci. Usually it is assumed that the number of clusters K is known and the problem is to find the best parameterization of each cluster model; which is readable as sending the model to at least one user computer for use thereof by the user computer on original data) (see col. 1, lines 25-38).

As per claims 21, Fayyad substantially teaches a method as claimed, wherein the user computer uses the model on original data to render a classification, and then sends the classification to the Web site (thus, each of the points in figure 4B may be thought of as a "guess" for the possible location of a mode in the underlying distribution the estimates are fairly varied but they exhibit "expected" behavior the subsampling produces a good separation between the two clusters; which is readable wherein the user computer uses the model on original data to render a classification, and then sends the classification to the Web site) (see col. 6, lines 13-15).

As per claim 22, Fayyad substantially teaches a method as claimed, wherein the model is sent to the user computer as a JAVA applet (see col. 1, lines 20-35).

*11) **Response to Arguments***

The Examiner will address the issues raised by the appellant in the order in which they appear in the appeal brief.

As per claims 1, 7 and 13, Appellant argues that the references do not teach or suggest:

In response to appellant's argument page 3 of the brief, "Fayyad does not teach maintaining the privacy of a user of the computer." However, Tendick teaches a method of preserving the privacy of individual records in a statistical database, see page 47, lines 1-2. Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Fayyad and Tendick with steps of maintaining the privacy of a user of the computer. This modification would allow the teachings of Fayyad and Tendick to improve the performance of the system and architecture for privacy preserving data mining, and provide optimal security, and a certain specified level of statistical accuracy for legitimate users of the database (see pages 60 and 61, lines 28 and 5-7).

On page 3 of the brief, Appellant stated that "explain where in the prior art all of the claimed limitations are taught or suggested." As previously stated in the final rejection mailed on July 15, 2002, the Examiner has drawn a mapping correspondence as indicated to the section I above.

In response to Appellant's argument on page 3 of the brief, Appellant should be note that obviousness is based upon improper hindsight reasoning, but it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

In response to Appellant's arguments on pages 3 and 4 of the brief, that "broad conclusory statements regarding the teaching of multiple references, standing alone, are not evidence," the test for combining references is not what the individual references themselves suggest but rather what the combination of the disclosures taken as a whole would suggest to one of ordinary skill in the art. In re Mclaughlin, 170 USPQ 209 (CCPA 1971).

In response to appellant's argument on page 4 of the brief, that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988)and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, Fayyad strongly suggests steps of the data clustering is important in a variety of fields including data mining, statistical data analysis, data compression, and vector quantization, see col. 1, lines 12-14.

On page 5 of the brief, Appellant stated that Fayyad does not use or suggest "using perturbed values of the original values at all." Examiner disagrees because Fayyad includes a starting point involves an additional analysis of the multiple solutions, the multiple clustering solutions from the initial data samples are each chosen as a starting point for clustering of all candidate solutions, see col. 3, lines 32-41. Further, in column 2, lines 20 through 22 and 27 through 30, Fayyad teaches the algorithm is deterministic and the solution is determined by the choice of an initial or starting point; and it has been well known that clustering algorithms are extremely sensitive to initial conditions. This implication discloses the use of perturbed values of the original values.

In response to appellant's arguments on page 5 of the brief, that "Fayyad nowhere considers privacy," against the references individually, one cannot show nonobviousness by

attacking references individually where the rejections are based on combinations of references. See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.*, 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986).

In response to applicant's argument on page 6, a prima facie case of obvious of is established when the teachings from the prior art itself would appear to have suggested the claimed subject matter to a person of ordinary skill in the art. Once such a case is established, it is incumbent upon appellant to go forward with objective evidence of unobviousness. In re Fielder, 471 F.2d 640, 176 USPQ 300 (CCPA 1973).

Examiner is entitled to give claim limitations their broadest reasonable interpretation in light of the specification.

Interpretation of Claims-Broadest Reasonable Interpretation

During patent examination, the pending claims must be 'given the broadest reasonable interpretation consistent with the specification.' Applicant always has the opportunity to amend the claims during prosecussion and broad interpretation by the examiner reduces the possibility that the claim, once issued, will be interpreted more broadly than is justified. In re Prater, 162 USPQ 541,550-51 (CCPA 1969).

Respectfully submitted,

Jean Bolte Fleurantin

Jean M. Corrielus
Art Unit 2172

JBF/

Hosain Alam
AU 2172